

The Use of Machine Learning Algorithms and Statistical Models to Classify Aphasia Severity

BY

Tea Kristiane Espeland Uggen

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR
THE DEGREE OF

DOCTOR OF PHILOSOPHY

UNDER THE SUPERVISOR OF

A./Prof. Tapan Rai, A./Prof. Erin Godecke and
Prof. Louise Ryan

UNIVERSITY OF TECHNOLOGY SYDNEY

SCHOOL OF MATHEMATICAL AND PHYSICAL SCIENCES

SEPTEMBER 2020

Certificate of Authorship

I, Tea Kristiane Espeland Uggen, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Science, at the University of Technology Sydney (UTS).

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution. This research is supported by an Australian Government Research Training program.

Production Note:

Signature: Signature removed prior to publication.

Date: 03/09 /2020

Acknowledgements

I wish to extend my deepest appreciation for my primary supervisor, A./Prof. Tapan Rai, for the opportunity to work on the most interesting and rewarding research project, for all your guidance and for the substantial amount of feedback you have given me over the past four years. I can see a distinctive improvement in my way of thinking, writing and presenting as a researcher from four years ago and this is largely thanks to you and the faith you have placed in me. Thank you for taking me on as your student and for always pushing me forward. I would also like to express my gratitude to my co-supervisors, A./Prof. Erin Godecke and Prof. Louise Ryan, for supporting me in my research, sharing your knowledge and providing me with numerous opportunities to share my research, collaborate and grow as a researcher.

A Ph.D. is no easy job, especially not without a supportive network of friends and colleagues. Thankfully I was lucky enough to hit the goldmine of friends in my fellow Ph.D. and postdoc colleagues. My most heartfelt thanks go out to our UTS statistics group for your friendship, your help with R and LaTeX bugs, your assistance in presentation preparations and last but not most importantly your emotional support during the downs of the emotional rollercoaster that is a Ph.D. A special shout out to Marianne, Hon, Luca, James, Wilson, Gioia, Bojana and Mahrita for always being there for me.

I want to thank the mathematics and statistics staff at UTS as well for taking the time to provide me with feedback throughout my degrees at UTS and for being such a knowledgeable and approaching team. I am very fortunate to have been surrounded by such great teachers and mentors. An especially big thanks to Dr. Stephen Bush and Prof. Matt Wand. Thank you as well to the helpful people at UTS Hacky Hour for helping me understand and work through my coding bugs the numerous times I've been stuck on a problem - I would not have made it through without your help.

I am also deeply grateful for my Australian family, uncle Erik, aunt Barbara and Ross, for your valuable input, encouragement and support during my studies.

Finally, I want to express my deepest thanks to my family in Norway, who have supported me during my nine years of studying in Sydney. In particular - my wonderful mom and dad for their constant support during my studies, for following my progress for every subject completed and for assisting as my non-technical audience for countless essays, presentations and articles. Furthermore, to the most encouraging and caring brother and sister in the world - thanks for listening to my Ph.D. rants and for all your words of encouragement. I also want to thank my two grandmothers for being the best, toughest and most inspiring role-models. Only the drive and courage passed on to me from you both could have gotten me through these past nine years. Thanks for all the knowledge and life experiences you have shared with me.

Abstract

Aphasia is a communication impairment usually caused by stroke, identified in approximately one third of stroke survivors. Determining the severity level for patients with aphasia is a necessary pre-cursor in determining the optimal rehabilitation pathway for each patient. Language assessments are used to evaluate aphasia severity. There are numerous existing language assessments in the aphasia literature. In this thesis, we will focus on two common assessments: the Western Aphasia Battery-Aphasia Quotient (WAB-AQ) and Discourse Analysis (DA). The WAB-AQ is considered a gold standard of aphasia assessments. However, it is lengthy and strenuous to undertake. DA is a shorter and more manageable assessment. However, there are no existing standards to classify aphasia severity from this assessment. Our research aims to determine whether it is possible to use the DA assessment to predict the aphasia severity level of the WAB-AQ. There are three research questions in this thesis: 1) Is it possible to use machine learning (ML) and natural language processing (NLP) algorithms to automatically identify the levels of aphasia severity from patients' speech transcripts, recorded during DA assessments? This objective is achieved by developing an NLP algorithm to convert the speech transcripts to numerical measures of speech; then performing classification using these measures to predict WAB-AQ severity level with various ML techniques. We extend the binary confusion matrix to a multi-class confusion matrix and develop associated model assessment metrics, specifically developed to tackle our three-class classification problem, to evaluate the predictive power of our models. We also determine the measures of speech from the NLP algorithm which are most important in classifying aphasia severity by developing an accuracy-based feature-selection algorithm able to incorporate multiple machine learning techniques. 2) How does the performance of our proposed method compare to a comparative baseline method? This baseline method uses existing clinical measures of DA as predictors in classification models to predict WAB-AQ severity level. 3) How do existing measures of DA compare among each other in their ability to predict aphasia severity? We use the measures individually as model predictors and compare the model assessment metrics from each model. Our results show that our proposed approach can be used to identify aphasia severity in post-stroke patients instantly with up to 73.6% accuracy. Furthermore, our findings indicate that our proposed method is superior in predicting aphasia severity level than the baseline method, with the NLP models yielding consistently higher model accuracies than the baseline models.

Keywords: *aphasia, classification, machine learning, natural language processing, severity*

Contents

1	Introduction: Stroke and Aphasia	1
1.1	Outcome Measures in Clinical Trials	1
1.2	Area of Application: Post-Stroke Aphasia	1
1.2.1	Types of Aphasia	2
1.2.2	Current Measures of Aphasia	3
1.2.3	The Western Aphasia Battery Aphasia Quotient	3
1.2.4	Discourse Analysis	5
1.3	Existing Aphasia Literature	7
1.3.1	Variability in Aphasia	7
1.3.2	Standardisation of Outcome Measures	8
1.3.3	Analysis of Language Impairment	8
1.4	Thesis Structure	10
2	Aims of this Thesis	11
2.1	Limitations of Existing Measures of Speech for Assessing Aphasia	11
2.1.1	Current Measures of Aphasia	11
2.1.2	Automated Algorithms for Measuring Speech	12
2.2	Research Objectives	12
2.2.1	Objective 1: Automatically Predict WAB-AQ Severity from DA Assessments	13
2.2.2	Objective 2: Comparing our Proposed Method to a Baseline Method	13
2.2.3	Objective 3: Comparison of Existing DA Measures in Predicting Aphasia Severity	14
2.3	The Subsequent Chapter	14
3	Data, Preparation and Analysis	15
3.1	Transcripts	15
3.2	Data Cleaning and Preparation	16
3.2.1	Speech Transcripts	16
3.2.2	Existing DA Measures	17
3.2.3	Future Transcription of DA Recording	17
3.3	Consistent Aphasic Speech Transcription (CAST)	18
3.4	Analysis: A Two-Step Statistical Approach	19
4	Step 1: Natural Language Processing for Text-Based Analysis	21
4.1	The Proposed NLP Algorithm	21
4.1.1	Linguistic Terminology	22
4.1.2	NLP Algorithm Development	24
4.2	De-Bugging	30
4.3	Algorithm in Pseudo-Code	32
5	Step 2: Classification of Aphasia Severity	34

5.1	Pre-classification of extreme cases	34
5.2	Predictors	35
5.3	Comparative Baseline Model	35
5.4	Machine Learning Techniques	36
5.4.1	Random Forest	36
5.4.2	Support Vector Machines	37
5.4.3	Ordinal Logistic Regression	37
5.5	Feature Selection	38
5.5.1	Motivation behind the development of CAFSA Algorithm	38
5.5.2	Algorithm in Pseudo-Code	41
6	Assessing Model Performance	42
6.1	Model Accuracy	42
6.2	The Binary Confusion Matrix	44
6.3	The Three-Class Ordinal Confusion Matrix	47
6.3.1	High and Low Cost	47
6.3.2	Extremal Cost	48
6.3.3	Model Performance Metrics	49
6.4	The K -Class Ordinal Confusion Matrix	50
6.4.1	Matrix Measures	51
6.4.2	Model Performance Metrics	51
6.5	The K -Class Nominal Confusion Matrix	53
6.6	Weighted Cohen's Kappa	54
6.7	Chapter Summary and Moving Forward	54
7	Exploratory Data Analysis	55
7.1	Proposed Method Variables	55
7.1.1	Descriptive Statistics of NLP Data	56
7.1.2	Barcharts of NLP variables	59
7.1.3	Correlation matrix of NLP variables	61
7.1.4	Boxplots of NLP variables across severity	61
7.2	Baseline Method	66
7.2.1	Assessment of descriptive statistics of DA variables	66
7.2.2	Correlation matrix of DA variables	67
7.2.3	Boxplots of DA variables across severity levels	68
7.2.4	Summary of the exploratory data analysis	68
8	Results	70
8.1	Proposed method	70
8.1.1	Naive Classification Model Output	70
8.1.2	Feature-Selection for Naive Classification	76
8.1.3	Selected Classifiers Model Output	76
8.2	Baseline (DA) method	78
8.2.1	Naive Analysis	78
8.2.2	Selected Classifiers Output	78

8.3	Proposed Method versus Baseline Method	81
9	Discussion	83
9.1	Objective 1: Developing an algorithm to automatically classify aphasia severity	83
9.1.1	Development of an NLP Algorithm to Convert Speech Transcripts to Numerical Data	84
9.1.2	Classification with the NLP-Derived Features of Speech	84
9.1.3	Determining the Superior Classification Model	85
9.1.4	Assessment of Feature Importance	86
9.2	Objective 2: Comparison of Proposed Method to a Baseline Method . . .	86
9.3	Objective 3: Comparison of Existing DA Measures in Classifying Aphasia Severity	88
9.4	The Following Chapter	89
10	Conclusions	90
10.1	Summary	90
10.2	Research Contributions	91
10.3	Research Limitations	92
10.4	Further Research Objectives	94
10.5	Concluding Remarks	96
A	Python	100
A.1	Part-of-Speech Tags	100
A.2	Programming Code	100
B	RStudio	111
C	Literature Search: Web of Science	118
C.1	Accuracy-based feature selection for multiple classification techniques . .	118
C.2	Multi-class confusion matrix	119

List of Figures

1	Examples of pictures used during picture description task	5
2	An example illustration of a given test observation (highlighted in orange) in a Random Forest for an outcome with three classes	37
3	Illustration of a CAFSA output: matrix of accuracies	39
4	Binary confusion matrix	45
5	Extension from binary to three-level ordinal confusion matrix	48
6	Complete three-level ordinal confusion matrix	49
7	Stacked barplots of means and medians of NLP variables	60
8	Correlation matrix of NLP variables	62
9	Boxplots of variables in the descriptive category	63
10	Boxplots of variables in the disfluency category	64
11	Boxplots of variables in the syntax category	65
12	Boxplots of variables in the relevance and similarity categories	66
13	Correlation matrix of DA variables	67
14	Boxplots of DA variables across WAB severity levels	68
15	Random Forest classification model output	72
16	Support Vector Machines (linear kernel) classification model output . . .	73
17	Support Vector Machines (radial kernel) classification model output . . .	73
18	Support Vector Machines (polynomial kernel) classification model output	73
19	Support Vector Machines (sigmoid kernel) classification model output . .	74
20	Ordinal Logistic Regression classification model output	74
21	K-Nearest Neighbours classification model output	74
22	Multinomial Logisitic Regression classification model output	75
23	Naive Bayes classification model output	75
24	Linear Discriminant Analysis classification model putput	75
25	Legend for variable importance across machine learning techniques	76
26	Visualisation of CAFSA-based variable importance	77
27	Naive DA analysis for Random Forest	79
28	Naive DA analysis for Support Vector Machines (linear)	79
29	Naive DA analysis for Support Vector Machines (radial)	79
30	Naive DA analysis for Support Vector Machines (polynomial)	79
31	Naive DA analysis for Support Vector Machines (sigmoid)	79
32	Naive DA analysis for Ordinal Logistic Regression	80
33	Naive DA analysis for <i>K</i> -Nearest Neighbours	80
34	Naive DA analysis for Multinomial Logistic Regression	80
35	Naive DA analysis for Naive Bayes	80
36	Naive DA analysis for Linear Discriminant Analysis	80
37	Random Forest and Support Vector Machines classification model compar- ison of proposed vs. baseline methods	82

List of Tables

1	Examples of transcripts of differing severity levels from picture description	16
2	Descriptive statistics of descriptive variables	57
3	Descriptive statistics of disfluency variables	57
4	Descriptive statistics of syntax variables	58
5	Descriptive statistics of relevance and similarity variables	59
6	Descriptive statistics of DA variables	67
7	MLT and measures of model performance for proposed models	78
8	MLT and measures of model performance for baseline models	81
9	Examples of unreliable output from the calculations of %CIU per minute	89
10	Part-of-speech tags available in the NLTK module	101